PSet_2

Paulino Diaz

February 5, 2020

Part 1

1.

If we wished to identify the effect of the number of children on how much a woman works, in an experimental world, we would be able to randomly assign different number of children to a significant sample of women, and regress this against their hours-worked-per-week. The coefficient of the variable *number of children* would give us the average treatment effect (ATE) of one more child on hours-worked-per-week. The random assignment of different number of children would ensure our estimate is unbiased. This assumes, however, that the effect is homogenous and constant across different total number of children. It also assumes that everyone is a complier.

If we instead believe the ATE changes depending on the total number of children, we might want to create different subsets of the experimental sample for different total number of children. The subsets would include combinations of women with n and n-1 total children, and compare their change in hours worked. Repeating the above regression would give us a LATE for each of the different subsets.

Additionally, if we believe there are non-compliers, the above regression would give us the ITT effect - which we could then weight against the probability of being a complier to get the ATT. If the effects are homogenous between compliers and non-compliers, the ATT = ATE.

2

- The before-after estimator would compare hours-worked-per-week for each woman, before and after getting one more child.
- The naive estimator would assume there is perfect compliance when there isn't - comparing the treatment and control groups without weighting for compliers - giving us an ITT effect instead of an ATE.
- We could have selection bias on unobservables, if the non-compliers are different from the compliers in ways we can't observe, and we could also have ommitted variable bias, if the compliers and non-compliers differ in things like gender, age, income level, etc.

3a

The estimated treatment parameter is the LATE, and it applies to the individuals in the sample whose fertility has been affected by having two same-sex children.

3b

- Compliers are the women who have a third child, but only because their first two children are of the same-sex.

- Defiers are the women who would have had a third child, but they didn't because their first two children are of the same sex.
- Always-takers are the women who have a third child, regardless of the sex of their previous two children.
- Never-takers are the women who would never have a third child, regardless of the sex of their previous two children.

3c

The exclusion restriction requires that having two same-sex children does not affect hours-worked-per-week, except through the increased likelihood of having a third child. That is to say, that the only link between hours-worked-per-week and the sex of the first two children is through the likelihood of having a third child. This implies that having two-same-sex children is completely random and unrelated to other observed or unobserved factors that determine fertility and labor force participation of the mother.

If the dependent variable is married women's husbands' income, then the exclusion restriction could be violated if we think the sex of the child affects the father's commitment to the family or changes the way father's treat their children.

3e

The monotonicity assumption implies that the effect of the instrument is always weakly increasing. This means that women who have two same-sex children are only more or equally likely to have one more child (none are less likely to have one more child). In other words, there are no defiers in the sample.

4

The study provides valuable information on the effect on labor supply for women who go from two to more than two children, and can help policymakers decide the magnitude of the subsidy for this group. However, these results are LATE, and not representative of the impact of childbearing in general since the sample is restricted to women going from two to more than two children and in a relatively young age-group. Older women with larger family sizes could require a very different intervention.

5

```
fertility_data <- read_dta(here("children-data.dta"))

clean_dta <- fertility_data %>%
  select(-childid) %>%
  unite(boy_twins, boy, twins) %>%
  pivot_wider(names_from = birthno, values_from = boy_twins) %>%
  separate("1", into = c("birth1", "twin1")) %>%
  separate("2", into = c("birth2", "twin2")) %>%
  select(-twin1) %>%
  mutate(twopluskids = case_when(nrchildren >= 3 ~ 1,
                                 nrchildren == 2 ~ 0),
         samesex = case_when(birth1 == birth2 ~ 1,
                             birth1 != birth2 ~ 0),
         nodiploma = case_when(educm < 12 ~ 1,
                               educm >= 12 ~ 0))
```

6

```
##
## Simple OLS with Robust SE
## ================================================================
##                            Hours-Worked-Per-Week of Mother
## ----------------------------------------------------------------
## More Than Two Children                -3.59***
##                                        (0.14)
##
## Constant                              19.01***
##                                        (0.09)
##
## N                                      65,664
## R2                                      0.01
## Adjusted R2                             0.01
## Residual Std. Error              18.10 (df = 65662)
## F Statistic                   616.10*** (df = 1; 65662)
## ================================================================
## Notes:                 ***Significant at the 1 percent level.
##                         **Significant at the 5 percent level.
##                         *Significant at the 10 percent level.
```

The results of the above regression estimate that having more than two children reduces a mother's average hours-worked-per-week by 3.5 hours. However, we should be skeptical of doing a causal interpretation of these results as it is likely that a mother's fertility and labor supply are jointly determined.

7

```
##
## OLS with Controls and Robust SE
## ================================================================
##                            Hours-Worked-Per-Week of Mother
## ----------------------------------------------------------------
## More Than Two Children                -3.84***
##                                        (0.15)
##
## Mother's Year of Birth                -0.31***
##                                        (0.02)
##
## No Highschool Diploma                  -0.22
##                                        (0.19)
##
## Has College Degree                    -1.46***
##                                        (0.22)
##
## Constant                              34.53***
##                                        (1.05)
##
## N                                      65,664
## R2                                      0.01
## Adjusted R2                             0.01
## Residual Std. Error              18.07 (df = 65659)
```

```
## F Statistic                  214.91*** (df = 4; 65659)
## ============================================================
## Notes:                 ***Significant at the 1 percent level.
##                         **Significant at the 5 percent level.
##                         *Significant at the 10 percent level.
```

The results of the above regression show that adding controls increases the magnitude of the effect of having more than two children, reducing a mother's hours-worked-per-week, on average, by 3.8 hours. The relationship between education and hours-worked-per-week also appears to be negative for both mother's without a highschool diploma and mother's who have a college diploma. However, the effect for no diploma is not statistically significant. Considering we do not observe the effect for those who have a highschool diploma, a possible story for explaining this negative relationship could be that not earning a highschool diploma decreases your labor prospects which negatively impacts your labor supply. Additionally earning a college degree increases your productivity, labor prospects and wages, allowing you to earn more while working less - thus also reducing your overall labor supply. Such a scenario would require that the unobserved effect of a highschool diploma is positive, increasing your labor prospects compared to those without a highschool diploma - while still requiring you to work more than those with a college diploma in order to earn a living wage.

Again, however, there are still reasons to believe that our variable of interest, $MoreThanTwoChildren$, and our outcome are jointly determined - which doesn't allow us to make a causal inference.

8a

```
##
## First Stage Regression: Two Same-Sex Children
## ============================================================
##                                  More Than Two Kids
## ------------------------------------------------------------
## Same-sex                              0.06***
##                                       (0.004)
##
## Mother's Year of Birth                -0.02***
##                                       (0.001)
##
## No Highschool Diploma                 0.18***
##                                       (0.005)
##
## Has College Degree                    -0.11***
##                                        (0.01)
##
## Constant                              1.25***
##                                        (0.03)
##
## N                                      65,664
## R2                                      0.04
## Adjusted R2                             0.04
## Residual Std. Error                0.48 (df = 65659)
## F Statistic                   721.82*** (df = 4; 65659)
## ============================================================
## Notes:                 ***Significant at the 1 percent level.
##                         **Significant at the 5 percent level.
##                         *Significant at the 10 percent level.
```

The above first stage regression shows that our instrument, *samesex*, is positively related with the probability of having more than two children - increasing this average probability by 6 percent. Our estimate of this effect is both strong and statistically significant, thus satisfying the instrument condition. If we also believe that having two same-sex children is a truly random occurrence, then the sex mix of a family is uncorrelated with a mother's labor supply (except through fertility) and our exclusion restriction is also satsified, making this a valid instrument.

8b

```
##
## 2SLS: Two Same-Sex Children
## ================================================================
##                            Hours-Worked-Per-Week of Mother
## ----------------------------------------------------------------
## More Than Two Children                -4.88**
##                                        (2.21)
##
## Mother's Year of Birth                -0.33***
##                                        (0.05)
##
## No Highschool Diploma                  -0.03
##                                        (0.44)
##
## Has College Degree                    -1.57***
##                                        (0.33)
##
## Constant                              35.87***
##                                        (3.02)
##
## N                                      65,664
## R2                                      0.01
## Adjusted R2                             0.01
## Residual Std. Error              18.08 (df = 65659)
## ================================================================
## Notes:                  ***Significant at the 1 percent level.
##                          **Significant at the 5 percent level.
##                          *Significant at the 10 percent level.
```

The treatment parameter for the above regression gives us the LATE of having more than two children when the first two children have the same sex. This gives us the effect for mothers with more than two children, whose fertility has been affected by having two same-sex children. All else equal, having more than two kids reduces this group's average hours-worked-per-week by 4.8 hours. Given we have a valid instrument, we can interpret these results causally, *but only for this group.*

9a

First stage:

$$TwoPlusKids_i = \pi_0 + \pi_1 * Twins_i + \pi_2 * YearOfBirthM_i + \pi_3 * NoDiploma_i + \pi_4 * CollegeM_i + \nu_i$$

Second stage:

$$HoursWorkedM_i = \beta_0 + \delta * \hat{TwoPlusKids_i} + \beta_2 * YearOfBirthM_i + \beta_3 * NoDiploma_i + \beta_4 * CollegeM_i + \epsilon_i$$

9b

```
##
## First Stage Regression: Twins
## ===============================================================
##                                  More Than Two Kids
## ---------------------------------------------------------------
## Twins                                  0.61***
##                                        (0.004)
##
## Mother's Year of Birth                 -0.02***
##                                        (0.001)
##
## No Highschool Diploma                  0.18***
##                                        (0.005)
##
## Has College Degree                     -0.11***
##                                         (0.01)
##
## Constant                               1.27***
##                                         (0.03)
##
## N                                       65,664
## R2                                       0.05
## Adjusted R2                              0.05
## Residual Std. Error              0.48 (df = 65659)
## F Statistic                  887.24*** (df = 4; 65659)
## ===============================================================
## Notes:                 ***Significant at the 1 percent level.
##                         **Significant at the 5 percent level.
##                         *Significant at the 10 percent level.
```

The above first stage regression shows that our instrument, *Twins*, is positively related to the probability of having more than two children - this is of course obviously true given that having twins in the second birth always means that the mother has more than two children. This satisfies our instrument condition. However, using twins as an instrument might fail to satisfy the exclusion restriction of our instrument, if we think that twinning probabilities appear to be correlated with some observed characteristics of the mother. The validity of the instrument will depend on whether we can control for these characteristics.

9c

```
##
## 2SLS: Twins
## ===============================================================
##                            Hours-Worked-Per-Week of Mother
## ---------------------------------------------------------------
## More Than Two Children                  -1.52
##                                         (1.24)
##
## Mother's Year of Birth                 -0.27***
##                                         (0.03)
##
## No Highschool Diploma                  -0.63**
```

```
##                                        (0.29)
##
## Has College Degree                    -1.20***
##                                        (0.26)
##
## Constant                              31.56***
##                                        (1.90)
##
## N                                      65,664
## R2                                      0.01
## Adjusted R2                             0.01
## Residual Std. Error              18.11 (df = 65659)
## ================================================================
## Notes:                  ***Significant at the 1 percent level.
##                          **Significant at the 5 percent level.
##                          *Significant at the 10 percent level.
```

The treatment parameter for the above regression gives us the LATE of having more than two children due to a twin in the second birth. This effect is for all mothers who have had more than two children due to twinning. According to the results, having more than two kids due to twinning reduces this group's average hours-worked-per-week by 1.5 hours - however, this result is not statistically significant. Even if it were, the magnitude of the effect is much smaller than for women who have a third child due to the gender of their first two children. A possible explanation for this, presented in the Angrist paper, could be the difference in ages between third childs born out of twinning and those that aren't. This assumes that the effect of children on the labor supply is different depending on their age.